**Assignment 1 – Introduction to Data Science**

**Exercise 1**

The average FEV1 values for smokers and non-smokers are: (3.2768615384615383, 2.5661426146010187).

The values show a lower average FEV1 for non-smokers which should indicates a decrease in lung function. This doesn’t match with what was expected: the FEV1 should be lower for the smokers group as it’s widely known that smoking affects the breathing.

**Exercise 2**

****

The plot shows that the FEV1 values are higher for the smokers group that for the non-smokers. The red line represents the median of the distribution. In the non-smokers data, some cross-hatches with high values of FEV1 are plotted even though the median results lower than the smokers median.

**Exercise 3**

The t value is: 7.1496081295

Degrees of freedom: 83.0

The p value computed manually is: 3.11735739253e-10

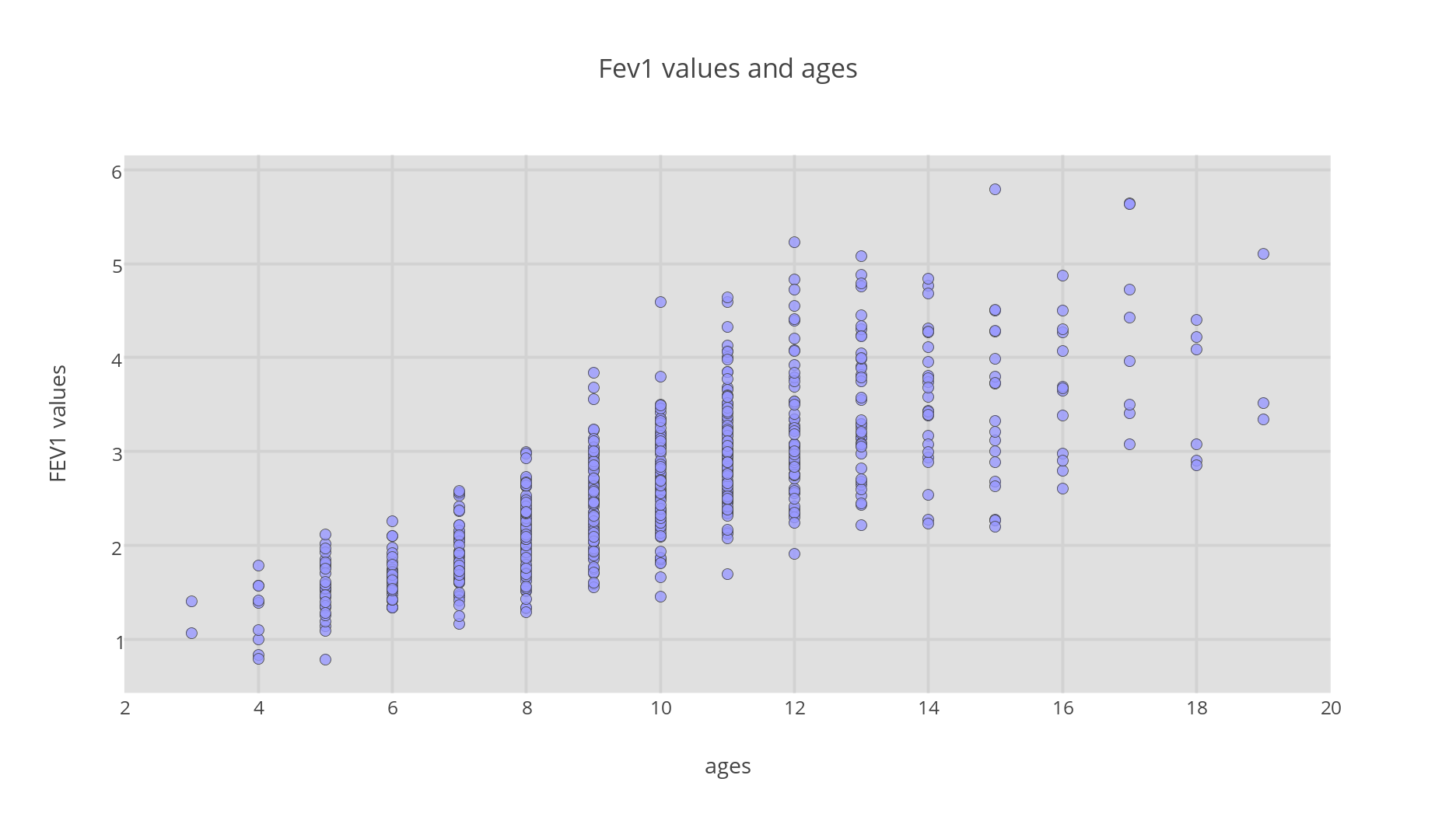
The p value computed with the built-in formula is: 3.07381274488e-10

The null hypothesis is that the two populations have the same mean. The hypothesis is rejected because the p-value resulting from the T-test is smaller than the significance level of α = 0.05.

The result is not surprising and

#TODO short discussion of the result

**Exercise 4**

****

The correlation coefficient manually computed is: 0.757617426421

The Pearson correlation coefficient computed with the built-in formula is: 0.75645898999

#TODO: say something about it

**Exercise 5**

